The evolution of MHC diversity: Evidence of intralocus gene conversion and recombination in a single-locus system

Bahr, Angela ; Wilson, Anthony B

Abstract: Gene conversion, the unidirectional exchange of genetic material between homologous sequences, is thought to strongly influence patterns of genetic diversity. The high diversity of major histocompatibility complex (MHC) genes in many species is thought to reflect a long history of gene conversion events both within and among loci. Theoretical work suggests that intra- and interlocus gene conversion leave characteristic signatures of nucleotide diversity, but empirical studies of MHC variation have rarely been able to analyze the effects of conversion events in isolation, due to the presence of multiple gene copies in most species. The potbellied seahorse (Hippocampus abdominalis), a species with a single copy of the MH class II beta-chain gene (MHIIb), provides an ideal system in which to explore predictions on the effects of intralocus gene conversion on patterns of genetic diversity. The genetic diversity of the MHIIb peptide binding region (PBR) is high in the seahorse, similar to other vertebrate species. In contrast, the remainder of the gene shows a total absence of synonymous variation and low levels of intronic sequence diversity, concentrated in 3 short repetitive regions and 1-12 SNPs per intron. The distribution of substitutions across the gene results in a patchwork pattern of shared polymorphism between otherwise divergent sequences. The pattern of nucleotide diversity observed in the seahorse MHIIb gene is congruent with theoretical expectations for intralocus gene conversion, indicating that this evolutionary mechanism has played an important role in MHC gene evolution, contributing to both the high diversity in the PBR and the low diversity outside this region. Neutral variation at this locus may be further reduced due to biases in nucleotide composition and functional constraints.

DOI: https://doi.org/10.1016/j.gene.2012.01.017

Manuscript revision (GENE-D-11-00641): *Gene* – 16/01/2012

The evolution of MHC diversity: Evidence of intralocus gene conversion and recombination in a single-locus system

Angela Bahr[1], Anthony B. Wilson*

Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland; [1]angela.bahr@uzh.ch

*Correspondence to:

Anthony B. Wilson

Institute of Evolutionary Biology and Environmental Studies,

University of Zurich,

Winterthurerstrasse 190,

8057 Zurich,

Switzerland

Tel: +41 44 635 4790

Fax: +41 44 635 4780

tony.wilson@ieu.uzh.ch

Abstract

Gene conversion, the unidirectional exchange of genetic material between homologous sequences, is thought to strongly influence patterns of genetic diversity. The high diversity of major histocompatibility complex (MHC) genes in many species is thought to reflect a long history of gene conversion events both within and among loci. Theoretical work suggests that intra- and interlocus gene conversion leave characteristic signatures of nucleotide diversity, but empirical studies of MHC variation have rarely been able to analyze the effects of conversion events in isolation, due to the presence of multiple gene copies in most species. The potbellied seahorse (*Hippocampus abdominalis*), a species with a single copy of the MH class II beta-chain gene (MHIIb), provides an ideal system in which to explore predictions on the effects of intralocus gene conversion on patterns of genetic diversity. The genetic diversity of the MHIIb peptide binding region (PBR) is high in the seahorse, similar to other vertebrate species. In contrast, the remainder of the gene shows a total absence of synonymous variation and low levels of intronic sequence diversity, concentrated in 3 short repetitive regions and 1-12 SNPs per intron. The distribution of substitutions across the gene results in a patchwork pattern of shared polymorphism between otherwise divergent sequences. The pattern of nucleotide diversity observed in the seahorse MHIIb gene is congruent with theoretical expectations for intralocus gene conversion, indicating that this evolutionary mechanism has played an important role in MHC gene evolution, contributing to both the high diversity in the PBR and the low diversity outside this region. Neutral variation at this locus may be further reduced due to biases in nucleotide composition and functional constraints.

Keywords:

Gene conversion; major histocompatibility complex; molecular evolution; *Hippocampus abdominalis*

Abbreviations:

MHC / MH, major histocompatibility (complex); MHIIb, major histocompatibility class II beta-chain; PBR, peptide binding region; PBS, peptide binding sites

1. Introduction

The major histocompatibility complex (MHC / MH in Actinopterygii), a key component of the adaptive immune system, is one of the most diverse gene families in the vertebrate genome (Janeway et al., 2002). The diversity of this gene complex is thought to be generated by a combination of mutation, gene duplication and loss, and recombination, and is maintained over time by selection (Ohta, 1991; Piertney & Oliver, 2006; Yeager & Hughes, 1999). The MHC peptide binding region (PBR) encodes a pocket in the molecule that allows the binding of specific pathogen-derived antigens, and this region typically exhibits the highest sequence polymorphism within the gene (Janeway et al., 2002).

Different regions of MHC genes are thought to be influenced by contrasting types of selection. The high diversity of the PBR, for example, is maintained by strong balancing selection (Ohta, 1991; Sommer, 2005; Spurgin & Richardson, 2010) and variation at nearby neutral sites may also be maintained by genetic hitchhiking (Hughes, 2000). Sexual selection is also thought to contribute to PBR diversity through disassortative mating, in which PBR-dissimilar individuals are preferred during mate choice (Milinski, 2006; Penn & Potts, 1999). In contrast, the region outside the PBR, critical for the secondary structure of the protein, is subject to purifying selection, with low levels of nonsynonymous substitution (Hughes & Nei, 1989).

Mutations create the diversity on which selection can act, but polymorphism can be further increased by recombination, bidirectional sequence exchange in which both alleles can be altered. Gene conversion, the unidirectional exchange of genetic material between homologous sequences of single or multiple loci, is thought to play an important role in the molecular evolution of gene families such as MHC (Chen et al., 2007; Hughes, 2000; Ohta, 1982; 2010; but see Martinsohn et al., 1999). Gene conversion events are thought to be responsible for both the high diversity of many gene families (e.g.

AB0 blood group locus, HLA class II region) (reviewed in Chen et al., 2007) and the patterns of concerted evolution found in many multigene families and highly repeated DNA sequences (e.g. histone genes, immunoglobulin genes, ribosomal RNA genes) (Ohta, 2010; Rada et al., 1990; Reusch & Langefors, 2005; reviewed in Nei & Rooney, 2005).

Both inter- and intralocus gene conversion are expected to increase the number of alleles at MHC loci via the non-reciprocal exchange of genetic material (Ohta, 1997; 1999). In the presence of positive selection, even over short evolutionary time scales, neutral variation located away from selected sites may be lost, resulting in dN > dS > *d* at loci experiencing gene conversion (Ohta, 1997; 1998; 1999). As gene conversion can involve the exchange of relatively short tracts of genetic material, both types of conversion may lead to a patchwork pattern of nucleotide polymorphism, clusters of shared mutations between otherwise divergent sequences (Martinsohn et al., 1999; Ohta, 2000; Parham et al., 1995).

Interlocus gene conversion is expected to lead to increases in the number of heterozygous sites and the frequency of synonymous (dS), non-synonymous (dN) and intronic (*d*) substitutions, as new mutants are introduced from divergent loci. Interlocus conversion consequently leads to low dN:dS ratios (Ohta, 1998; 1999), and may ultimately lead to the concerted evolution of homologous loci (Ohta, 1982). Gene homogenization may occur if variable regions are repeatedly exchanged during conversion events (Martinsohn et al., 1999), or if the converted sequences experience directional selection (Hughes, 2000).

In contrast, intralocus gene conversion simply reshuffles existing variation within a locus, eroding existing nucleotide diversity (Ohta, 1999). While high levels of intralocus conversion can effectively eliminate variation at neutral sites, overdominant selection limits the loss of functional variation, and genetic hitchhiking maintains synonymous variation tightly linked to selected sites. As such, the

interaction of selection and intralocus conversion can lead to exceptionally high dN:dS and dS:d ratios (Ohta, 1999).

The effect of gene conversion on MHC polymorphism is still under debate (Martinsohn et al., 1999; Nei & Rooney, 2005) and several alternative explanations for the nucleotide pattern observed in MHC genes have been suggested. The high MHC allele number and genetic diversity of these genes may be maintained by positive selection alone (Ohta, 1991), or by maternal-fetal interactions, the selective disadvantage of allele sharing between mother and offspring (Hedrick & Black, 1997; Ohta, 1998). Other studies have supported a birth-and-death model of MHC evolution, involving a combination of gene duplication and loss and overdominant selection on the PBR of the gene (e.g. Hughes & Nei, 1989). Nevertheless, simulation studies have shown that overdominant selection alone is insufficient to create a dN:dS ratio larger than one (Ohta, 1991; 1997), which requires positive selection at least in the short term (Ohta, 1998). Shared polymorphism may result from co-ancestry or convergence (Klein & Figueroa, 1986; Kriener et al., 2000), but these mechanisms are unlikely to explain a patchwork pattern of genetic variation in the absence of mutation rates higher than those observed at MHC genes or unrealistically long time periods (>20 Myr ago for a shared motif of 3 amino acids) (Reusch & Langefors, 2005; Yeager & Hughes, 1999).

Studies on threespined sticklebacks, the bony fish for which MH genes have been best characterized, provide evidence that gene duplication and gene conversion have been important mechanisms driving MH allelic diversity in species with multiple copies of these genes (Reusch et al., 2004; Reusch & Langefors, 2005; Schaschl & Wegner, 2007). The effects of inter- and intralocus gene conversion and recombination can, however, not be differentiated in multi-locus systems, unless MH alleles can be unambiguously assigned to specific loci (e.g. Reusch et al., 2004), a task which has proven difficult in the MH system (e.g. Michel et al., 2009; Reusch & Langefors, 2005; Sato et al., 1998).

Surprisingly, despite the intense analysis of MHCIIb genes in vertebrates (reviewed in Piertney & Oliver, 2006; Wegner, 2008), full-length gDNA sequences are extremely rare, particularly in fishes. Studies investigating intronic MHIIb diversity have either included full-length sequences of only a subset of loci present in the investigated species (Reusch et al., 2004; Sultmann et al., 1994), or have been restricted to the introns adjacent to the PBR (Michel et al., 2009; Reusch & Langefors, 2005). Complete MHIIb sequences covering the full genetic diversity present in a species would provide an important tool to compare the molecular evolution of these genes between different groups of vertebrates. To this end, we have studied the evolution of the full-length MH class II beta-chain gene (MHIIb) of the potbellied seahorse, *Hippocampus abdominalis*, a species with a single locus inherited in a Mendelian fashion (Bahr & Wilson, 2011).

2. Materials and Methods

2.1. Samples

Complete MHIIb gene sequences were obtained from 10 *H. abdominalis* specimens in order to characterize the distribution of sequence polymorphism within the gene. Two samples were collected from Sydney Harbor in 2003 and 3 individuals were collected from 3 Tasmanian locations in 2003 and 2004 (Wilson & Martin-Smith, 2007). Mainland Australian and Tasmanian seahorses are genetically distinct (Wilson, A.B., unpublished data). The remaining 5 samples originate from several generations of an aquaculture population, which was founded with seahorses collected from several Tasmanian sites (for details on this population see Bahr & Wilson, 2011). Previous analyses of a large sample of this population (N = 101 individuals) revealed high genetic diversity in exon 2, the region encoding the MHIIb PBR, a pattern similar to that observed in other vertebrate species (Bahr & Wilson, 2011). The high variability at neutral microsatellite loci observed in *H. abdominalis* (Bahr & Wilson, 2011; Wilson & Martin-Smith, 2007) suggests that recent population bottlenecks or founder effects are unlikely to have influenced gene diversities in this species (Kim et al., 1999).

105

106 **2.2. Full-length MHIIb gene sequencing**

107 Genomic DNA was extracted from muscle tissue using a standard proteinase K / phenol-chloroform
108 protocol (Bruford et al., 1998). To characterize the full-length MHIIb gene of the seahorse, primers
109 were designed based on the complete MHIIb sequence of a single individual (Hiab-DAB*01/02, Bahr
110 & Wilson, 2011). Sequences were aligned using BioEdit v.7.0.9.1 (Hall, 1999) and primers were
111 designed using Primer3 v.0.4.0 (Rozen & Skaletsky, 2000). Primers used additionally to those
112 previously published (Bahr & Wilson, 2011) are provided in Table 1.

113

114 To obtain full-length MHIIb sequences, primers MHIIb-E1F2 and MHIIb-E6R were used under the
115 long-range PCR conditions described in Bahr and Wilson (2011). These primers amplify the region
116 between exons 1 to 6 of the gene. 5' and 3' ends of sequences were completed using the primer
117 combinations MHIIb-UTR5F with MHIIb-I2R4 and MHIIb-E5F with MHIIb-UTR3R under the same
118 PCR conditions. Full-length PCR products were first sequenced directly, producing heterozygous
119 sequences with degenerate positions labeled according to IUPAC nomenclature. Second, products
120 were cloned prior to sequencing, using a Topo TA Cloning Kit (Invitrogen) following the
121 manufacturers' recommendations. 2-8 positive colonies per individual were picked into 25 µL of
122 ddH_2O, directly PCR amplified and sequenced in a polymorphic region of the gene (exon 2 resp.
123 introns 1 and 5) to identify allelic variants. For each allele, 1-2 cloned products were sequenced to
124 completion. Cloned products were compared to direct sequences in order to detect nucleotide
125 misincorporations and to infer allelic phase. Up to 15 different combinations of nested primers were
126 used to amplify shorter products of the seahorse MHIIb gene (300-2000 bp in length) to confirm
127 homozygosity in individuals putatively homozygous for the full gene sequence. Samples were
128 prepared for sequencing as described elsewhere (Bahr & Wilson, 2011).

129

PCR reactions spanning three intronic single-bp repeat regions (dA and dT in intron 2, dC in intron 4) preferentially amplified one of the two alleles in 5 individuals, irrespective of PCR product length. Such a pattern may be caused by large differences in repeat length, which may result in the preferential amplification of the shorter allele (Pompanon et al., 2005). The addition of either Betaine or 5% DMSO to PCR reactions in order to reduce secondary structure formation did not recover the second allele in these individuals. Primers flanking these repeat regions were designed, and consecutive PCRs, excluding those single-bp repeats, allowed for the detection of heterozygous sequences in these individuals (see Suppl. Figure S1).

2.3. Processing of sequences

Sequence data were assembled using Sequencing Analysis 5.2 (Applied Biosystems), aligned with Muscle v.4.0 (Edgar, 2004), and verified by eye in BioEdit v.7.0.9 (Hall, 1999). Analyses on full-length exon sequences were conducted on uncropped sequences of 249 codons (exons 1-6, 747 bp), corresponding to 12 distinct amino acid sequences. The complete dataset of exonic and intronic MHIIb sequences yielded 20 distinct alleles in our sample (see Suppl. Figure S1) but, as single-bp repeat regions in intronic sequences could not be covered in all individuals (see above), the final dataset contains full length alleles from 14 of 20 sequences (encompassing 14 different alleles, total length ≤ 3525 bp). In many cases, sequences identical in exon 2 show intronic variability, suggesting either the evolutionary conservation of functional alleles with neutral variation accumulating over time, or the convergent evolution of exon 2 alleles. Full-length MHIIb sequences were assigned composite allele names, reflecting both exon 2 and full-length identity. The alleles Hiab-DAB*0101 and *0102, for example, have exon 2 sequences identical to the previously published Hiab-DAB*01 allele (Bahr & Wilson, 2011), but differ in intronic sequence and are numbered consecutively.

2.4. Analyses of sequence polymorphism

DnaSP v.4.90.1 (Librado & Rozas, 2009) was used to calculate standard diversity estimates. A sliding-window analysis of synonymous and non-synonymous variation of concatenated exon sequences (length 30, steps 12) was produced in SWAAP v.1.0.3 (Pride, 2000). Intron diversity *d* (the average number of differences per site for intron sequences) was calculated using Mega v.4.0.2 (Tamura et al., 2007) under a Jukes-Cantor model. CENSOR was used for detection of transposons (Kohany et al., 2006) in full-length DNA sequences and a CpG-island search was performed using EMBOSS CpGPlot/CpGReport (www.ebi.ac.uk/Tools/emboss/cpgplot/) (window size: 100 bp, step size: 1 bp, island length: 200 bp).

2.5. Positive selection

Mega v.4.0.2 was used to calculate dN and dS, as well as to test for positive selection in the exon dataset, applying a Z-test under a Jukes-Cantor model (10,000 permutations). Peptide binding sites (PBS) were inferred through homology modeling to human HLA-DRB*01 alleles (Brown et al., 1993; Reche & Reinherz, 2003). Site-specific positive selection was inferred using Codeml, implemented in the PAML v.4.2b package (Yang, 2007).

2.6. Recombination / Gene conversion

We tested for recombination and gene conversion in full-length MHIIb gene sequences using the default settings of RECCO v.0.93 (1,000 permutations) (Maydt & Lengauer, 2006). RECCO reconstructs each sequence from alignment data under various scenarios of recombination and mutation, and the best fit model is inferred using a permutation test.

3. Results

MHIIb gene sequences were obtained for 10 seahorses collected from five populations (Suppl. Figure S1, GenBank ID: JN858015-32). The MHIIb gene of the seahorse consists of six exons (see Bahr &

Wilson, 2011 for gene map), with a total length of 3512-3525 bp and a total exon length of 747 bp. In high-quality DNA samples, we were able to amplify the complete gene using primers MHIIb-E1F and MHIIb-E6R. In several individuals, intronic repetitive elements inhibited amplification of single alleles, necessitating the sequencing of shorter gene fragments. In these animals, the complete MHIIb sequence was obtained for one of the two alleles, while sequence data for the second allele near the repetitive regions of the introns are lacking (see 2.2 and 2.3, Suppl. Figure S1). MHIIb sequences comprise 20 different alleles due to intron variability and differences in repeat lengths, 14 of which provided complete, unambiguous sequences which were used to investigate intron diversity.

3.1. Sequence polymorphism in the PBR

The 9 MHIIb exon 2 sequences recovered here (Hiab-DAB*01, *04, *05, *08, *09, *11, *13, *16, and *17) are a subset of the 17 alleles identified in a larger sample of 101 individuals in a previous study (Bahr & Wilson, 2011), and include 25 polymorphic nucleotide sites and 17 amino acid differences. Only 2 of the 25 variable sites identified are synonymous (dN = 0.050, dS = 0.012), leading to a strong signal of positive selection on exon 2 (Z-Test p = 0.010). The nucleotide diversity π of the seahorse exon 2 in this dataset is 0.040. Variable sites were shared among captive-bred individuals and samples from natural populations.

3.2. Sequence polymorphism outside the PBR

The 20 MHIIb alleles found in the seahorse include 12 distinct exonic alleles. Polymorphism outside the region encoding the PBR of the seahorse is extremely low, with only two non-synonymous variants (located in exons 3 and 5) and a complete absence of synonymous substitutions (Fig. 1). A strong signal of positive selection over all exons is detected when the PBR is included (Z-Test, p = 0.003, Table 2), a signal which is absent when PBS are omitted from the analysis (Z-Test, p = 0.208, Table 2), despite the fact that the amino-acid substitution in exon 5 appears to be under site-specific positive selection (PAML p = 0.001). The exonic nucleotide diversity π of the seahorse MHIIb gene is

0.015, with an average of 11 nucleotide differences between sequences. The amino acid sequence alignment of the complete MHIIb gene consists of 19 polymorphic sites and 12 distinct alleles (data not shown).

Complete intron sequences are available for 14 / 20 full-length MHIIb alleles. Length differences between these alleles (≤ 13 bp) reflect variation in repeat regions and single bp indels (Fig. 2, Suppl. Figure S1). Intron variability in the seahorse MHIIb gene is low (Fig. 3), with 1 – 12 variable nucleotide sites per intron for an indel-free alignment. The nucleotide diversity over all introns (p-distance ± SE = 0.004 ± 0.001, Table 2) is equivalent to the synonymous substitution rate over all exons (dS ± SE = 0.004 ± 0.003), but 3 times lower than dS in exon 2 (dS ± SE = 0.012 ± 0.010). Intron regions within 14 bp of exons (mean = 69 bp, range 14 – 177 bp) are completely conserved across all alleles.

Exons and introns differ in GC content. Exons are GC-biased (0.45 AT vs. 0.55 GC), whereas introns have a strong AT bias (0.61 AT vs. 0.39 GC). This difference in nucleotide composition between exons and introns is statistically significant (Mann-Whitney U-Test: $n_1 = 12$, $n_2 = 14$, $U = 168$, $p < 0.001$). A high AT content in introns might, at least in part, be explained by the presence of AT-rich transposable elements (Chamary et al., 2006), and MHC genes are prone to the accumulation of transposons due to their high heterozygosity (van Oosterhout, 2009). Fragments of 5 DNA-transposons were detected in the MHIIb sequences of *H. abdominalis* (transposon class: 4x hAT, 1x Mariner/Tc1), all of which have an AT content >50% (data not shown).

A comparison of nucleotide variability in 14 full-length MHIIb alleles shows the conservation of short sequence tracts between otherwise divergent sequences, leading to a patchwork pattern of nucleotide variation (Fig. 2). This pattern could be influenced by the significant levels of intralocus recombination inferred in our dataset (Recco p = 0.013), with an estimated number of 5 recombinant

MHIIb full-length alleles. Recombination breakpoints were detected in the 5' region of exon 2, the 5' to middle region of intron 2 and the 5' region of intron 4. Exon 3 of the seahorse MHIIb gene was detected as CpG-island (ratio observed : expected CG = 0.93), regions typically associated with high levels of recombination and/or gene conversion (Högstrand & Böhme, 1999).

4. Discussion

The high variability of the peptide-binding region of the seahorse MHIIb gene, the region interacting with antigens, has been created and maintained by a combination of positive selection and intralocus recombination, consistent with findings from other vertebrates (see also Bahr & Wilson, 2011). Strikingly, the remainder of the gene shows a complete absence of synonymous variation and extremely low levels of intronic diversity. Furthermore, short sequence tracts are conserved between otherwise divergent sequences of the seahorse MHIIb gene (Fig. 2), as expected under intralocus gene conversion.

4.1. Nucleotide diversity

Despite similarities between the level of non-synonymous variation in the region encoding the PBR of the seahorse MHIIb and that of other teleosts, the nucleotide diversity of the full-length seahorse MHIIb gene is low ($\pi = 0.015$ over all exons incl. PBS, $d = 0.004$ over all introns; *Salmo salar*: $\pi = 0.026$ (Stet et al., 2002); *Leiocassis longirostris*: $\pi = 0.056$ (Shen et al., 2011); *Gasterosteus aculeatus*: $d = 0.016$ for intron 2 (Reusch & Langefors, 2005); *Perca fluviatilis*: $d = 0.071$ for intron 1 (Michel et al., 2009)). We detected only two exonic substitutions (both nonsynonymous) outside of exon 2. The absence of synonymous substitutions outside this region contrasts with previous studies, in which purifying selection has been shown to eliminate recessive deleterious mutations and contribute to an excess of synonymous relative to nonsynonymous substitutions outside the PBR (Hughes & Nei, 1989). Unfortunately, almost all studies to date have focused exclusively on the peptide binding region

of the teleost MH, and given the paucity of previous studies investigating full-length MH genes in multiple individuals, it is unclear whether the low level of synonymous variation in the seahorse MHIIb gene reflects a unique evolutionary history in this group or a broader evolutionary pattern. A comparable study of coding DNA in *Salmo salar*, a species with a single MHIIb gene, revealed only 7 alleles in 84 specimens, with 1 synonymous and 2 non-synonymous substitutions outside exon 2 (Stet et al., 2002), a pattern qualitatively similar to that recovered here. In contrast to this low diversity, Chinese longsnout catfish (*Leiocassis longirostris*), another species with a single gene copy, showed 7 synonymous and 11 nonsynonymous substitutions outside exon 2 in 6 complete cDNA alleles (n = 5 specimens) (Shen et al., 2011). Klein et al. (1993) obtained two complete cDNA sequences that differed by 22 substitutions (9 synonymous and 13 nonsynonymous) outside exon 2 from a single *Aulonocara hansbaenschi* individual, a species with at least 2 MHIIb loci.

A reduced intron diversity relative to the synonymous variation detected in exon 2 has been observed in other teleosts (Cyprinidae - intron 1 vs exon 2: Hughes, 2000; threespined sticklebacks - exon 2 vs intron 2: Reusch & Langefors, 2005; Eurasian perch - intron 1 vs exon 2: Michel et al., 2009). Consequently, the sharp contrast in the substitution patterns of the region encoding the PBR and the remainder of the gene appears to be common for MHC genes, with the potbellied seahorse reflecting the lower range of nucleotide diversity observed in other teleosts.

4.2. Molecular evolution

MHC genes have been suggested to evolve under a combination of mutation, selection, recombination and gene conversion, gene duplication and/or loss, and drift. These mechanisms are responsible for the maintenance of contrasting patterns of nucleotide variation among different regions in MHC genes. The seahorse MHIIb gene shows a high number of nonsynonymous substitutions in the region encoding the PBR, while genetic diversity outside this region is low, with variation distributed in a patchwork fashion across the gene (Fig. 2).

285

286 The high exon 2 diversity of the seahorse MHIIb gene is maintained by positive selection on peptide
287 binding sites, a pattern commonly observed in other studies (Wegner, 2008; Yeager & Hughes, 1999).
288 A significant signature of intralocus recombination and/or gene conversion was detected in the
289 seahorse MHIIb exon 2 (see also Bahr & Wilson, 2011), both powerful mechanisms which increase
290 MHC allele number and dN:dS ratio in combination with positive selection (Ohta, 1997; 1998; 1999).
291 A high dN:dS ratio may also be achieved by overdominant selection in combination with short term
292 selection (Hughes & Nei, 1989; Ohta, 1998). Analyses of exon 2 in a large sample of *H. abdominalis*
293 individuals found no evidence of either an excess of MHIIb heterozygous individuals (HWE Exact
294 Test: $p = 0.08$), or the non-random association of particular MHIIb alleles (see Bahr & Wilson, 2011),
295 suggesting that overdominant selection is unlikely to be the major force driving the high dN:dS ratio in
296 this species.

297

298 Nucleotide diversity outside the region encoding the PBR of the seahorse MHIIb gene is much lower
299 than that detected in exon 2, a pattern typical of MHC genes (Yeager & Hughes, 1999; Hughes, 2000).
300 However, in contrast to previous studies, the seahorse shows no synonymous substitutions outside this
301 region. Intronic diversity (*d*) is also lower in the seahorse than that found in other teleosts, and shared
302 polymorphism is found between otherwise divergent MHIIb alleles. The observed pattern of
303 nucleotide variability in the seahorse MHIIb gene is congruent with expectations for intralocus gene
304 conversion. Allelic gene conversion has been shown to reduce the number of heterozygous sites of
305 MHC genes (Ohta, 1997; 1998; 1999). This form of conversion is also expected to lead to an excess of
306 dS relative to *d*, as recombination across the exon-intron boundary homogenizes neutral MHC gene
307 sequences. Following gene conversion, synonymous mutations linked to positively selected sites are
308 maintained by hitchhiking, while sites outside the PBR are eliminated by random drift (Hughes, 2000;
309 Ohta, 1998; 1999; Reusch & Langefors, 2005). Intralocus gene conversion also has the potential to
310 produce the observed patchwork pattern of shared polymorphism between alleles (Ohta, 1999; 2000;

311 Parham et al., 1995). The detection of several recombination breakpoints in the seahorse MHIIb gene,
312 the lower nucleotide diversity in introns compared to exon 2, and the patchwork pattern of nucleotide
313 diversity all indicate that intralocus gene conversion has a major influence on the molecular evolution
314 of the seahorse MHIIb gene.
315
316 Evidence for the importance of recombination and gene conversion on MH genes has been found in
317 other teleost species (sticklebacks: Reusch et al., 2004, Reusch & Langefors, 2005; salmonids: Aguilar
318 & Garza, 2007, Langefors et al., 2001; perch: Michel et al., 2009), but due to difficulties in the
319 assignment of MH alleles to different loci, it has proven difficult to discriminate the relative
320 importance of intra- and interlocus sequence exchange in such systems. Despite the use of a large
321 number of primers in both conserved and variable regions of the MHIIb gene (Table 1), we found no
322 evidence of additional gene copies and/or pseudogenes in the potbellied seahorse. The fact that only a
323 single expressed copy of MHIIb was detected in a 454 sequencing screen (Bahr & Wilson, 2011)
324 supports the existence of a single functional copy of this gene in the seahorse. Despite the clear
325 evidence for a single extant locus of MHIIb in this species, interlocus conversion may have
326 contributed to the present-day genetic diversity at this locus via a birth-and-death model of MHC
327 evolution (Nei & Rooney, 2005).
328
329 The deficit of synonymous variation found in the seahorse MHIIb gene could also be influenced by
330 functional constraints on apparently neutral sites. Under this scenario, functionally important
331 mutations assumed to be silent would be deleterious and would be eliminated by purifying selection
332 (Chamary et al., 2006; Hughes, 2000). Such functional constraints could include a biased synonymous
333 codon usage to maximize translational efficiency, and/or to optimize mRNA stability, as well as
334 conserved exon-intron junctions involved in splicing control (Chamary et al., 2006; Hughes, 2000).
335 Deviations from base composition parity may also effectively reduce dS and *d* (Hughes, 2000; Wolfe
336 et al., 1989). The seahorse MHIIb gene fits all these patterns and its biased GC-content, both at 3rd

337 codon positions (76%) and in introns (39%), is consistent with expectations on functional constraints
338 and on the relationship of GC content and neutral variation (Hughes & Nei, 1989).
339
340 ## 5. Conclusions
341
342 Intralocus gene conversion is thought to be an important process shaping molecular diversity via the
343 exchange of genetic material among divergent alleles within genes. This mode of molecular evolution
344 is expected to leave a prominent signature on genes experiencing balancing selection, producing a
345 patchwork pattern of nucleotide substitutions and a heterogeneous distribution of neutral variation.
346 Congruent with theoretical expectations for intralocus recombination, the seahorse MHIIb gene shows
347 both of these characteristics. The high dN:dS ratio in the region encoding the MHIIb PBR of the
348 seahorse is typical of vertebrate MHC genes and has been influenced by a combination of positive
349 selection and intralocus gene conversion and recombination. While recombination has been shown to
350 effectively reduce intron diversity relative to exons in other species, this pattern may also be
351 influenced by a biased nucleotide composition in the seahorse.

Acknowledgements

We are grateful to K.N. Stölting for his suggestions for lab work and data analyses. We also thank P. Hedrick, L. Keller, T. Ohta and P. Reche for helpful discussions. Funding for this work has been provided by the University of Zurich.

Supplementary Material

Supplementary Figure S1:

Nucleotide alignment of complete MHIIb sequences for 10 *H. abdominalis* individuals.

References

Aguilar, A., Garza, J.C., 2007. Patterns of historical balancing selection on the salmonids major histocompatibility complex class II beta gene. J. Mol. Evol. 65, 34-43.

Bahr, A., Wilson, A., 2011. The impact of sex-role reversal on the diversity of the major histocompatibility complex: Insights from the seahorse (*Hippocampus abdominalis*). BMC Evol. Biol. 11, 121.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., Wiley, D.C., 1993. 3-Dimensional structure of the human class-II histocompatibility antigen HLA-DR1. Nature 364, 33-39.

Bruford, M.W., Hanotte, O., Brookfield, J.F.Y., Burke, T., 1998. Multi- and single locus DNA fingerprinting. in: Hoelzel, A.R. (Eds.), Molecular Analysis of Populations: A Practical Approach. IRL Press, Oxford, pp. 287-336.

Chamary, J.V., Parmley, J.L., Hurst, L.D., 2006. Hearing silence: Non-neutral evolution at synonymous sites in mammals. Nat. Rev. Genet. 7, 98-108.

Chen, J.M., Cooper, D.N., Chuzhanova, N., Ferec, C., Patrinos, G.P., 2007. Gene conversion: Mechanisms, evolution and human disease. Nat. Rev. Genet. 8, 762-775.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792-1797.

Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95-98.

Hedrick, P.W., Black, F.L., 1997. Random mating and selection in families against homozygotes for HLA in South Amerindians. Hereditas 127, 51-58.

Högstrand, K., Böhme, J., 1999. Gene conversion of major histocompatibility complex genes is associated with CpG-rich regions. Immunogenetics 49, 446-455.

Hughes, A.L., 2000. Evolution of introns and exons of class II major histocompatibility complex genes of vertebrates. Immunogenetics 51, 473-486.

Hughes, A.L., Nei, M., 1989. Nucleotide substitution at major histocompatibility complex class II loci - Evidence for overdominant selection. Proc. Natl. Acad. Sci. USA 86, 958-962.

Janeway, C.A., Travers, P., Walport, M., Shlomchik, M., 2002. Immunologie, fifth ed. Spektrum, Heidelberg, Berlin.

Kim, T.J., Parker, K.M., Hedrick, P.W., 1999. Major histocompatibility complex differentiation in Sacramento River chinook salmon. Genetics 151, 1115-1122.

Klein, D., Ono, H., O'hUigin, C., Vincek, V., Goldschmidt, T., Klein, J., 1993. Extensive MHC variability in cichlid fishes of Lake Malawi. Nature 364, 330-334.

Klein, J., Figueroa, F., 1986. Evolution of the major histocompatibility complex. CRC. Crit. Rev. Immunol. 6, 295-386.

Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7, 474.

Kriener, K., O'hUigin, C., Klein, J., 2000. Conversion or convergence? Introns of primate DRB genes tell the true story. Proceedings: Major Histocompatibility Complex - Evolution, Structure and Function 354-376.

Langefors, K., Lohm, J., von Schantz, T., 2001. Allelic polymorphism in MHC class IIb in four populations of Atlantic salmon (*Salmo salar*). Immunogenetics 53, 329-336.

Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451-1452.

Martinsohn, J.T., Sousa, A.B., Guethlein, L.A., Howard, J.C., 1999. The gene conversion hypothesis of MHC evolution: A review. Immunogenetics 50, 168-200.

Maydt, J., Lengauer, T., 2006. Recco: Recombination analysis using cost optimization. Bioinformatics 22, 1064-1071.

Michel, C., Bernatchez, L., Behrmann-Godel, J., 2009. Diversity and evolution of MHIIb genes in a non-model percid species - The Eurasian perch (*Perca fluviatilis* L.). Mol. Immunol. 46, 3399-3410.

Milinski, M., 2006. The major histocompatibility complex, sexual selection, and mate choice. Annu. Rev. Ecol. Evol. Syst. 37, 159-186.

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121-152.

Ohta, T., 1982. Allelic and nonallelic homology of a supergene family. P. Natl. Acad. Sci.-Biol. 79, 3251-3254.

Ohta, T., 1991. Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. P. Natl. Acad. Sci. USA 88, 6716-6720.

Ohta, T., 1997. Role of gene conversion in generating polymorphisms at major histocompatibility complex loci. Hereditas 127, 97-103.

Ohta, T., 1998. On the pattern of polymorphisms at major histocompatibility complex loci. J. Mol. Evol. 46, 633-638.

Ohta, T., 1999. Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. Immunol. Rev. 167, 319-325.

Ohta, T., 2000. An attempt to measure the patchwork pattern observed among alleles at major histocompatibility complex loci. J. Mol. Evol. 51, 21-25.

Ohta, T., 2010. Gene conversion and evolution of gene families: An overview. Genes 1, 349-356.

Parham, P., Adams, EJ., Arnett, K.L., 1995. The origins of HLA-A,B,C polymorphism. Immunol. Rev. 143, 141-180.

Penn, D.J., Potts, W.K., 1999. The evolution of mating preferences and major histocompatibility complex genes. Amer. Nat. 153, 145-164.

Piertney, S.B., Oliver, M.K., 2006. The evolutionary ecology of the major histocompatibility complex. Heredity 96, 7-21.

Pompanon, F., Bonin, A., Bellemain, E., Taberlet, P., 2005. Genotyping errors: Causes, consequences and solutions. Nat. Rev. Genet. 6, 847-859.

Pride, D.T., 2000. SWAAP Version 1.0.0 - Sliding windows alignment analysis program: A tool for analyzing patterns of substitutions and similarity in multiple alignments. Distributed by the author.

Rada, C., Lorenzi, R., Powis, S.J., Vandenbogaerde, J., Parham, P., Howard, J.C., 1990. Concerted evolution of class I genes in the major histocompatibility complex of murine rodents. P. Natl. Acad. Sci. USA 87, 2167-2171.

Reche, P.A., Reinherz, E.L., 2003. Sequence variability analysis of human class I and class II MHC molecules: Functional and structural correlates of amino acid polymorphisms. J. Mol. Biol. 331, 623-641.

Reusch, T.B.H., Langefors, A., 2005. Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. J. Mol. Evol. 61, 531-541.

Reusch, T.B.H., Schaschl, H., Wegner, K.M., 2004. Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. Immunogenetics 56, 427-437.

Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers, in: Krawetz, S., Misener, S. (Eds.), Bioinformatics methods and protocols: Methods in molecular biology. Humana Press, Totowa, NJ, pp. 365-386.

Sato, A., Figueroa, F., O'hUigin, C., 1998. Cloning of major histocompatibility complex (MHC) genes from threespine stickleback, *Gasterosteus aculeatus*. Mol. Mar. Biol. Biotechnol. 7, 221-231.

Schaschl, H., Wegner, K.M., 2007. Contrasting mode of evolution between the MHC class I genomic region and class II region in the three-spined stickleback (*Gasterosteus aculeatus* L.; Gasterosteidae : Teleostei). Immunogenetics 59, 295-304.

Shen, T., Xu, S.X., Yang, M., Pang, S.Y., Yang, G., 2011. Molecular cloning, expression pattern, and 3D structural analysis of the MHC class IIb gene in the Chinese longsnout catfish (*Leiocassis longirostris*). Vet. Immunol. Immunopathol.141, 33-45.

Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. 2, 16.

Spurgin, L.G., Richardson, D.S., 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. R. Soc. B 277, 979-988.

Stet, R.J.M., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B.P., Grimholt, U., 2002. Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. Immunogenetics 54, 320-331.

Sultmann, H., Mayer, W.E., Figueroa, F., O'hUigin, C., Klein, J., 1994. Organization of MHC class IIb genes in the zebrafish (*Brachydanio rerio*). Genomics 23, 1-14.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.

van Oosterhout, C., 2009. Transposons in the MHC: The Yin and Yang of the vertebrate immune system. Heredity 103, 190-191.

Wegner, K.M., 2008. Historical and contemporary selection of teleost MHC genes: Did we leave the past behind? J. Fish Biol. 73, 2110-2132.

Wilson, A.B., Martin-Smith, K.M., 2007. Genetic monogamy despite social promiscuity in the pot-bellied seahorse (*Hippocampus abdominalis*). Mol. Ecol. 16, 2345-2352.

Wolfe, K.H., Sharp, P.M., Li, W.H., 1989. Mutation rates differ among regions of the mammalian genome. Nature 337, 283-285.

Yang, Z.H., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586-1591.

Yeager, M., Hughes, A.L., 1999. Evolution of the mammalian MHC: Natural selection, recombination, and convergent evolution. Immunol. Rev. 167, 45-58.

Figure legends:

Figure 1: Distribution of synonymous (dS) and non-synonymous (dN) variation across exons of 12 MHIIb alleles of the seahorse. A sliding-window analysis with a window-length of 30 and a step-length of 12 was used.

Figure 2: Plot of nucleotide variability for 14 full-length *H. abdominalis* MHIIb alleles, showing the patchwork pattern of nucleotide variation at this locus. Exons and repetitive intron sequences are annotated.

Figure 3: Entropy plot of MHIIb variation in *H. abdominalis*, showing the distribution of polymorphic sites among 14 full-length alleles.

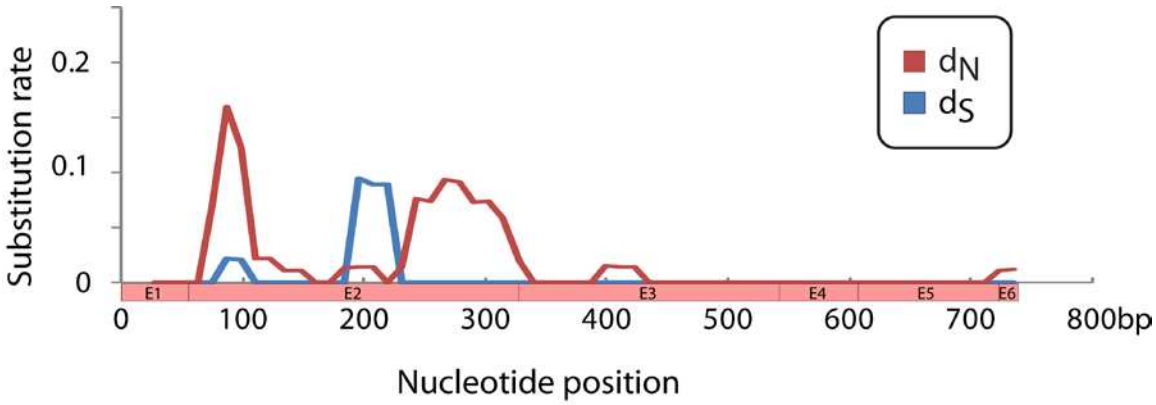
Table 1: Primers used to amplify and sequence MHIIb in *H. abdominalis.*

Name	**Sequence 5'-3'**	**Location**
MHIIb-UTR5F	CAAGTTTGTGCTCAGGTTGG	5'-UTR
MHIIb-E2F4	AACTCGAGTGACCAGAATGACATC	Exon 2
MHIIb-I2F	TAGGGCCTGACGAATATGGA	Intron 2
MHIIb-I2F4	AACGGAATCCATTTGGGAGT	Intron 2
MHIIb-I2R6	CAATGATTGTTCGGGTGTGA	Intron 2
MHIIb-E3F2	GCCTTACGTCAGACTTCACTCG	Exon 3
MHIIb-E3R3	GGCGTGTAGACCAGGTGAGA	Exon 3
MHIIb-E4F	GTGGAACACGCCAGCCTT	Exon 4
MHIIb-I4F2	TTTCCCACACGGTATCACAA	Intron 4
MHIIb-I4R	AACATCTCTCGGGTTGGTTG	Intron 4
MHIIb-I4R2	ACTTTACAGCAGGGGTCTTCA	Intron 4
MHIIb-E5F2	GCTGGACTGACTCTGGGTGT	Exon 5
MHIIb-UTR3R	ATCACTCAGTGCGAGCAGAA	3'-UTR

Table 2: Synonymous and non-synonymous substitution rates of full-length seahorse MHIIb alleles.

Locus	**Length**	**N**	**Alleles**	**dN**	**dS**	**dN / dS**	**d**
Exon 2	273	10	9	0.050	0.012	4.17*	
Exons 1-6	747	10	12	0.019	0.004	4.75*	
Exons 1-6 non-PBS	675	10	11	0.007	0.003	2.33^{ns}	
Introns	≤ 2775	10	14				0.004

Figure 1: Distribution of synonymous (dS) and non-synonymous (dN) variation across exons of 12 MHIIb alleles of the seahorse. A sliding-window analysis with a window-length of 30 and a step-length of 12 was used.



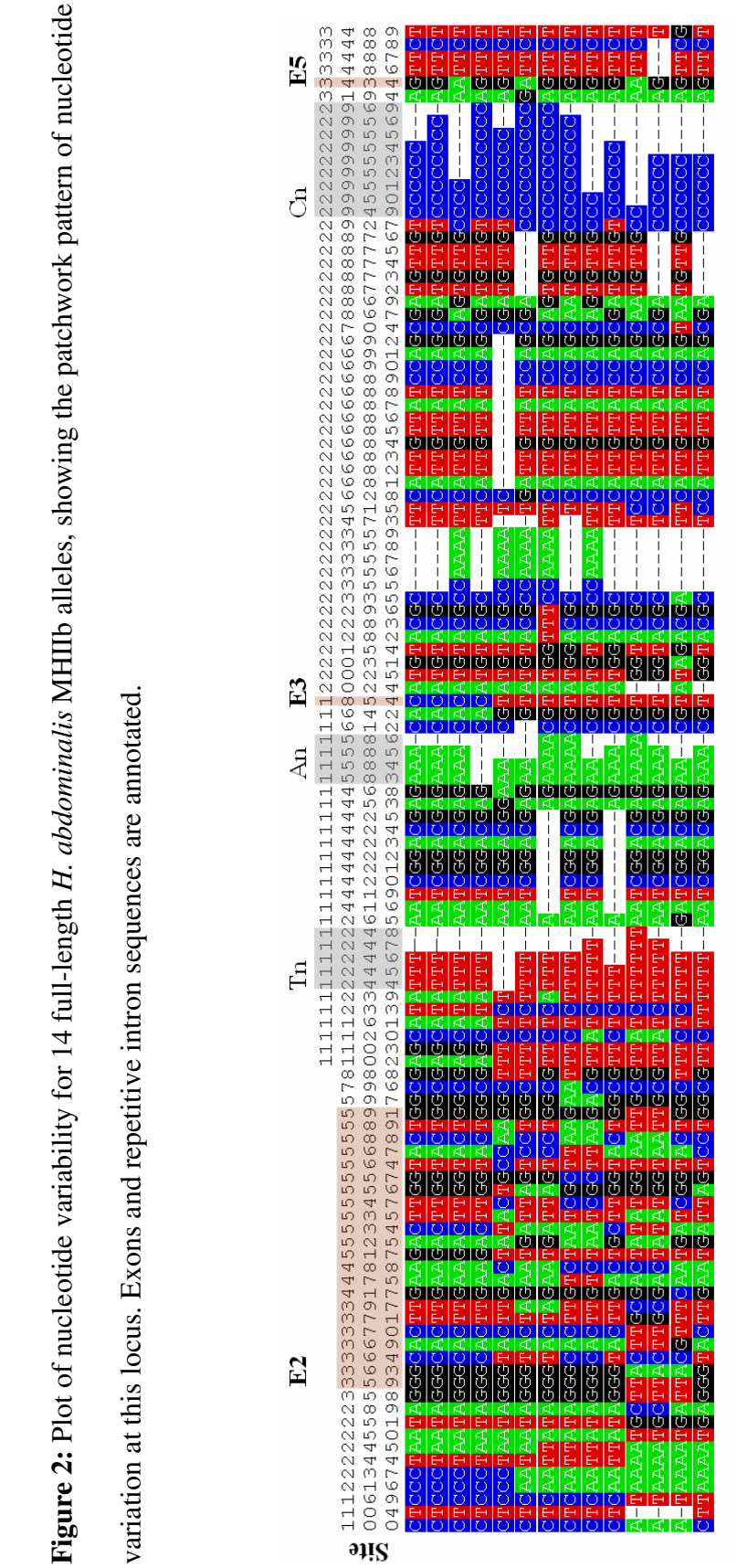


Figure 2: Plot of nucleotide variability for 14 full-length *H. abdominalis* MHIIb alleles, showing the patchwork pattern of nucleotide variation at this locus. Exons and repetitive intron sequences are annotated.

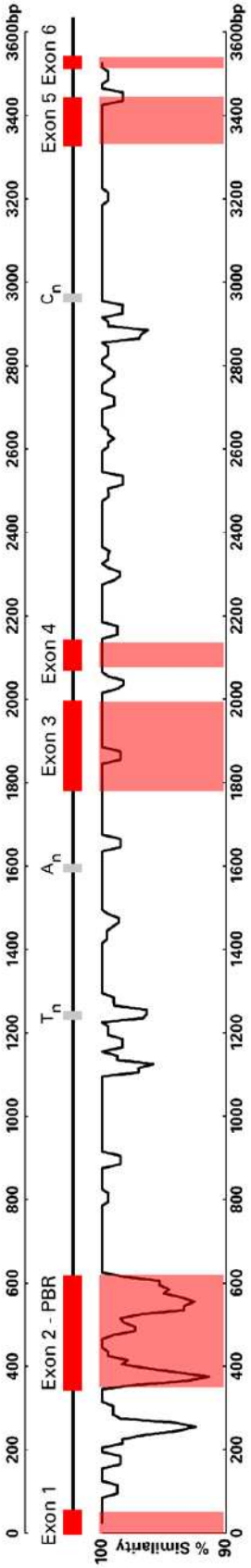


Figure 3: Entropy plot of MHIIb variation in *H. abdominalis*, showing the distribution of polymorphic sites among 14 full-length alleles.

Supplementary Figure S1: Alignment of 20 complete MHIIb nucleotide sequences. Dots indicate positions identical to the first sequence. Exons are shaded in grey and intronic repetitive regions are highlighted with clear boxes. Hiab-DAB*0401 corresponds to Hiab-DAB*01, and Hiab-DAB*0402 to Hiab-DAB*02 from Bahr & Wilson (2011).